DOCUMENT RESUME

ED 035 514 24 RE 002 346

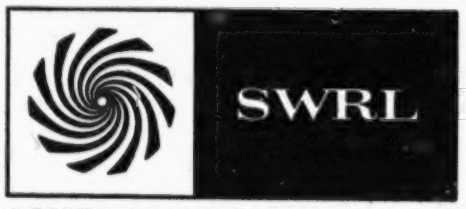
AUTHOR	Potter, Thomas C.
TITLE	A Taxonomy of Cloze Research, Part I: Readability and Reading Comprehension.
INSTITUTION	Southwest Regional Educational Lab., Inglewood, Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO	TR1
BUREAU NO	BR-6-2865
PUB DATE	1 Jun 68
CONTRACT	OEC-4-7-062865-3073
NOTE	51p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.65
DESCRIPTORS	*Cloze Procedure, Evaluation Methods, Language Research, *Measurement Techniques, Methodology, *Readability, Reading Comprehension, *Reading Research, *Research Tools, Test Construction, Test Reliability, Test Validity

ABSTRACT

An overall look is taken at the cloze research method as a new tool for measuring readability. The construction of a cloze readability test is described as well as several studies made on the reliability and validity of such tests. Included also are some methodological considerations to be kept in mind when constructing a cloze test. Data is reported from some cloze technique investigations, using children from first grade through high school and adults from a variety of populations, which suggest that the cloze technique is applicable to many types of communication and that it can be used to discriminate among the readability levels of passages and among the reading comprehension levels of readers. A description is given of the most valid and reliable cloze test for measuring passage difficulty as shown by research. It is suggested that more research must be done in several fundamental areas about this relatively new technique and that there are several new areas in which the technique may be used. A bibliography is included. (NH)

RE

BR 6-2865 TR-1
PA-24
OE/BR



SOUTHWEST REGIONAL LABORATORY FOR EDUCATIONAL RESEARCH & DEVELOPMENT

ED035514

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED

BY Southwest Regional Laboratory

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

RE002 346

A TAXONOMY OF CLOZE RESEARCH, PART I: READABILITY AND READING COMPREHENSION

TR 1 1JUNE, 1968

This publication was prepared pursuant to a
contract with the United States Department of
Health, Education and Welfare, Office of Education.

© 1968
SOUTHWEST REGIONAL LABORATORY for
Educational Research and Development
11300 La Cienega Boulevard
Inglewood, California, 90304

A TAXONOMY OF CLOZE RESEARCH, PART I: READABILITY AND READING COMPREHENSION

Thomas C. Potter

INTRODUCTION

The first comprehensive statement of the new research method, the cloze procedure, and its theoretical background were introduced in 1953 by Wilson L. Taylor in an article entitled "Cloze Procedure: A New Tool for Measuring Readability" in the Journalism Quarterly. During the past 15 years, research involving this new tool has appeared in a great variety of investigations.

At the heart of the procedure is a functional unit of measurement called the "cloze". Just as there is an apparent tendency to "see" a not quite complete circle as a whole circle by "mentally closing the gap" and making the image conform to a familiar shape, a mutilated sentence is filled in by completing those words that make the finished language pattern conform to the intended or apparently intended meaning. Thus, in the examples "chickens cackle and _____ quack", or "Merry Christmas and _____ New Year", the test-taker who writes in the correct word scores one cloze unit correctly by completing the sentence in accord with that language pattern. Taylor suggests that in order to satisfactorily complete a cloze, the subject must first know the meanings (the patterns or symbol-meaning

relationships), the forms (patterns of letters) of most or all of the words involved, and also the meanings of the given combinations of both in a particular sentence structure. In other words, one must guess what the mutilated sentence means as a whole and then complete its pattern accordingly. Cloze then, is defined by Taylor as a method for intercepting a message from a transmitter, mutilating its language pattern by deleting words, and administering it to receivers in such a way that their attempts to make the patterns whole again potentially yield a measure of their ability to deal with the general meaning and form intended.

The construction of a cloze readability test as outlined by Taylor, includes five steps: (a) delete a number of words selected by some random method without any regard for the function or the meaning of the specific words, (b) reproduce each mutilated passage with a blank space of some standard length in place of every missing word, (c) give the mutilated passage to representative samples of the subjects in question, (d) ask subjects to fill one word in each blank, determining from the context of the remaining words what that missing word should be, (e) total the correct number of responses for each passage and consider these totals readability scores. In contrasting the cloze totals of various passages, the one with the highest score would be considered most readable, the second one next most readable, and so on.

Obviously this procedure is similar to the familiar completion test in that the subject is instructed to fill in the blank spaces. It differs from the typical sentence completion test in that the deleted words are not pre-evaluated and selected according to the relative importance in the sentence. Obviously, the cloze procedure does not require an expert for either test construction or administration. The cloze deals with contextually interrelated series of blanks, not isolated ones. Rather than dealing directly with meaning, the cloze repeatedly samples the extent of likeness between the language patterns used by the communicator and that used by the receiver.

Initial inquiry into the nature of this procedure properly involved the testing of its validity in relation to readability. Taylor (1953, 1957) and numerous other investigators (Bormuth, 1962; Gallant, 1965; Ruddell, 1963) using only minor modifications in Taylor's method, have supported the initial findings that the cloze procedure ranks passages at differing difficulty levels in the same order as do older readability formulas.

Some of the principal intervening variables relating to the cloze procedure have come from Osgood's (1962) concept of "total language context". These include verbal factors such as grammatical skills, the effective use of multitudes of symbols, and non-verbal cues such as past experience and intelligence.

Osgood's theory of communication suggests that redundancies and

transitional probabilities lead to the development of dispositional mechanisms that play a large part in transmitting and receiving messages.

Redundancy. For example, "man coming" means the same as the redundant statement "a man is coming this way now". It is suggested that the latter is more like ordinary English; it indicates the singular number of the subject three times (by "a", "man", "is"), the present tense twice ("is coming" and "now"), and the direction of action twice ("coming" and "this way"). Such repetitions of meaning, it is suggested by Taylor, make it possible to replace "is", "this", "way", or "now" should they be deleted.

Transitional probabilities. Some words appear more frequently than others in certain patterns or sequences. For instances, the theory of transitional probability indicates that "Merry Christmas" is a more likely combination than "Merry Birthday". In like manner, "please pass the _____" is more often completed by "salt" than by "sodium chloride" or "blow torch". Some transitions from one word to the next are therefore more probable than others.

Habits of expression take over most of the work of translating an individual's meaning into an organized series of language symbols for transmission to others. Words appearing in sequences that fit the existing receiving habits of the reader can be understood with minimal effort. When the words appear in less familiar sequences, comprehension or understanding may be slower; a sufficiently improbable pattern may seem nonsensical.

The justification for the random deletion method used by Taylor is based on the fact that if enough words are deleted, the blanks will come to represent proportionately all kinds of words to the extent that they occur. Taylor suggests that cloze scores appear to be the measure of the aggregate influences of all factors which interact to effect the degree of correspondence between the language patterns of the transmitter and those of the receiver.

CONTEXTUAL CONSTRAINTS IN SENTENCE AND PARAGRAPH RECONSTRUCTION

Because the problem of contextual constraint is closely tied to that of the theoretical consideration of the cloze procedure, two studies on contextual constraint will be briefly discussed here. Aborn, Rubenstein, and Sterling (1959) suggested four hypotheses in relation to contextual constraint: (a) the predictability of words belonging to a given class is in general inversely related to the size of that class, (b) increasing the context beyond ten words between deletions does not seem to increase the accuracy of word prediction

(the length at which context contains maximum effectiveness lies between five and ten words), (c) a bilaterally distributed context exerts greater constraint than a totally preceding or a totally following context of the same length, (d) when the frequency of word class occurrence is more controlled, words have almost the same predictability in all positions in the sentence with the exception of final positions where predictability is much lower. MacGinitie (1961), indicated that Aborn, et al. (1959) raised the question of whether the constraints upon words in continuous discourse were cumulative. He explored the constraints within complete cloze paragraphs to determine the effect of differing distances between omissions. It was concluded that the influence of context upon a particular word choice in English prose decreased rapidly as the distance of the context from the word increased to five words. Beyond that point the distance between blanks seemed to have little effect on cloze scores.

METHODOLOGICAL CONSIDERATIONS

The cloze test is constructed by deleting words according to a predetermined system. The subject is given a set of instructions, generally very simple in nature, which include the following (Bormuth, 1964c):

1. Write only one word in each blank.
2. Try to fill in every blank. Don't be afraid to guess.
3. Wrong spelling will not count against you if we can tell what word you meant.

Subjects may be given a short practice exercise. To date, no empirical investigations involving children have considered the effect of differing test instructions on cloze scores. Tests are normally untimed and are administered in the classroom.

In scoring cloze tests, the exact word deleted is the criterion for correctness most often used. However, some designs call for the correct scoring of responses differing from the deleted word in semantic meaning or grammatical inflection. Taylor (1953) found that scores obtained by counting both grammatically correct synonyms (S.G.C.) and exact words (E.W.) deleted, were not significantly superior to scores obtained by using only the exact word criterion. Rankin (1957) and Ruddell (1963) found that S.G.C. plus E.W. scoring resulted in slightly increased variances on reading comprehension test scores. All scores obtained by counting grammatically correct responses exhibited positive correlations with each other. Bormuth (1965b) studied the correlations between standardized reading test scores and cloze test scores obtained by counting words as correct if their inflections were correct in context. Responses were further classified according to whether the stem of the response exactly matched, was synonymous with, or semantically unrelated to the deleted words. Multiple regression analyses of these data showed that scores based on responses which exactly matched the

deleted words in both inflections and word stems, accounted for 95 percent of the comprehension test variance that could be predicted from the total set of cloze test scores. It was concluded that the most economical and objective method of scoring cloze tests, the exact word method, showed the most valid results.

Most investigators have scored misspellings correct when the response was otherwise correct. No studies to date have investigated interscorer reliability in this practice. In like manner, the illegibility of written responses has not received consideration.

VALIDITY OF CLOZE TESTS

A substantial body of research has concerned itself with the validity of the cloze procedure when used to determine the readability of materials designed for children. While traditional readability formulas have, for many years, been applied to instructional materials to determine the suitability of reading difficulty in relation to the abilities of the students for whom the material was designed, the formulas currently available have validities that range from .5 to .7 (Bormuth, 1967a). Most readability formulas take into account only a small number of linguistic variables. Almost all formulas use traditional word frequencies as indicators of vocabulary difficulty. Some formulas also include sentence length or number of syllables per sentence. In any case, the variables taken into account are, in some respects, crude and outdated.

Cloze tests, as noted above, take into account many variables affecting passage readability. Taylor (1953) has compared the agreement of cloze score rankings of passages of varying difficulty and readability rankings of the same passages by two commonly-used formulas (Dale and Chall, 1948 and Flesch, 1949). It is reported that passages were rank-ordered similarly by each technique. The superiority of the cloze procedure was demonstrated when passages of obvious difficulty could be reasonably evaluated by the cloze procedure and not by the readability formulas.

Traditionally, reading comprehension ability is measured by allowing a subject to read a passage, and then measuring his knowledge of the content of that passage by his performance on multiple-choice items written for this purpose. Studies have reported significant correlation between cloze test scores and scores on specially-written multiple-choice comprehension tests (Bormuth, 1962; Jenkinson, 1957; Taylor, 1956).

A substantial number of investigations have reported high correlations between cloze tests and standardized tests of reading achievement. It is interesting to note that correlations with cloze scores are frequently higher for vocabulary measures than for comprehension measures. This is well illustrated by Bormuth (1967b, p. 9. See Table I).

Table I

Correlations Between Cloze Readability Tests and Standardized Tests of Reading Achievement

Study	Subjects	Tests	Correlations
Jenkinson (1957)	High School	Cooperative Reading C2	
		Vocabulary	.78
		Level of Comprehension	.73
Rankin (1957)	College	Diagnostic Survey	
		Story Comprehension	.29
		Vocabulary	.68
		Paragraph	.60
Fletcher (1959)	College	Cooperative Reading C2	
		Vocabulary	.63
		Level of Comprehension	.55
		Speed of Comprehension	.57
		Dvorak-Van Wagenen	
		Rate of Comprehension	.59
Hafner (1963)	College	Michigan Vocabulary Profile	.56
Ruddell (1963) (5 cloze tests)	Elementary	Stanford Achievement Paragraph Meaning	.61-.74
Weaver and Kingston (1963, 2 cloze tests)	College	Davis Reading	.25-.51
Green (1964)	College	Diagnostic Reading Survey Total Comprehension	.51

Unlike the standardized tests, cloze measures avoid the confounding effect of additional text materials such as comprehension questions and multiple-choice responses. Through the use of cloze tests, readability may be determined directly on the materials in question with the subject population for which it is intended. The 75 percent comprehension score has a long tradition of acceptance and widespread use in current practice (Betts, 1946; Harris, 1962; Thorndike, 1917). According to this criterion, a passage is determined suitable for a pupil's instructional use if he responds correctly to 75 percent or more of the items presented related to his understanding of that passage. In some investigations multiple-choice tests are constructed over the same passages used in cloze tests. In other investigations, cloze tests are constructed from passages which have been designed for use in standardized reading tests. In both cases cloze scores between 40 percent and 45 percent have been found comparable to the 75 percent criterion (Bormuth, 1967a).

READING COMPREHENSION AND INFORMATION GAIN

It is suggested by some investigators that reading comprehension is a generalized skill, best represented by the score obtained by finding the difference between pretest and posttest scores over a reading passage. Comprehension measured in this manner has been referred to by Coleman as information gain. Scores obtained in the traditional manner confound both knowledge acquired as a result of reading, the knowledge possessed before the reading of the passage and a memory factor. The usefulness of cloze as pretest and posttest measures of information gain has been considered by Coleman (1966). His findings indicated that bilateral constraint and/or the information given in cloze tests was such that little information was gained by reading the undeleted passage. The differences between means (information gain) for college students was found insignificant and the correlation between two cloze scores was .93. The few subjects used in this investigation and the undisclosed nature of the passages leaves the finding open to question. Another researcher (Taylor, 1956) has reported significant gain between pretest cloze scores and posttest cloze scores over a one week period.

RELIABILITY OF CLOZE TESTS

Several studies have investigated the reliability of cloze tests. It has been pointed out (Fletcher, 1959; Bormuth, 1962) that cloze tests frequently contain a number of very difficult and very easy items which are less efficient discriminators than items in the intermediate range. This fact, according to Bormuth (1962), may contribute to high correlations between cloze tests and other measures. It may also effect test-retest, and split-half reliability statistics. However, these same types of difficult and easy items in cloze tests may be an asset since they contribute to a test's validity with a variety of subjects over a wide range of difficulty levels. Skewed

distributions are infrequently reported when cloze tests are carefully administered.

SUMMARY OF INVESTIGATIONS INVOLVING THE CLOZE TECHNIQUE

Cloze tests have been used with children from the first grade level through high school. Subjects have also been drawn from adult populations such as military trainees and college students. The first section of this summary, Section A, is concerned with investigations of school-age subjects. Since age and reading ability were not considered as independent variables in studies using adult subjects, these investigations are grouped separately in Section B.

Section A--children

Two studies have involved primary (first, second, and third grade) children. Gallant (1965) considered the reliability and validity of cloze tests with first, second, and third grade children; the data suggest that cloze tests are appropriate for this age group (see Table II). A second part of the same investigation argued that sentence length may contribute to the variance in cloze scores for first and second graders. It must be noted that for first graders the cloze test was modified in such a way that each deletion became a three-option, multiple-choice test. In order to increase sentence lengths, unspecified modifications were made in a passage such that Spache Readability Measures were increased approximately one-half grade. The artificiality of the resulting prose may have confounded the comparisons made.

Deutsch (1964) investigated first and fifth grade subjects' cloze scores which were derived from deleted portions of teachers' classroom speech and from children's speech. The relationships of first and fifth grade scores to race, socio-economic status, and sex, were also investigated. While split-half reliability varied considerably, cloze scores were significantly related to I.Q. A main effect favoring high social status was found in addition to a social status by sex interaction (favoring High S.E.S., girls) on many scores. Only one instance of main effect involving race (Caucasian) was found. However, this difference disappeared when scores were covaried with I.Q. The auditory method of cloze test presentation resulted in significantly lower means than those derived from the written form. Deutsch states that the oral cloze test deserves emphasis since it more closely approximates the stimulus conditions of a classroom.

While many aspects of this study appear sound, the fact that cell sizes in this complex between-subjects design drop to three subjects, may be cause for concern in generalizing from the reported findings. Training instructions used in preparing subjects for cloze test responses were minimal. Little consideration has been given to the order effects of the various cloze instruments used, particularly with fifth grade subjects. It is possible that practice effects confounded

Table II. Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Part I Grade = 1st, 2nd, 3rd N = 273 Gallant, 1965	Part I D.R. every 5th word. Paragraph Reading Section, Metropolitan Reading Test. 1st grade material--made 3 option, multiple choice. Scoring--"reasonable approximation agreeing in person and tense."	Part I Metropolitan Achievement Tests for Grades 1 & 2. Metropolitan Reading Test Grades 3 & 4.	Subjects from two schools in Bedford Indiana. Part I Compare cloze score and standardized measures. Correlations tested for significance.	Part I Cloze scores reliable (split halves & S.B.) .90-.97. Standardized measures and cloze scores overall r .65-.81 (<.01) No sex differences.
Part II same as Part I	Part II Five passages from simplified folk tale from basal reader matched to S grade level. Set A: Spache readability 1.8, 1.9, 2.5, 2.9, 3.3 Set B: Sentence length was increased in unspecified manner to Spache level of 2.1, 2.4, 3.0, 3.7, 4.3. 1st grade material made into a 3 option multiple choice.	Part II No other measures used.	Part II Set A and B cloze scores compared using same subjects as Part I. Correlations tested for significance.	Part II Cloze score means higher for set A than set B (<.01) for first and second graders. No differences found for third graders.

other findings. The practice of covarying I.Q. test results with cloze test results may be valid when individually administered intelligence tests are used. However, group administered tests of any kind necessitate the use of verbal instructions, both written and oral, which may confound the measurement of listening and reading comprehension skills (see Table III).

An investigation by McLeod (1966) used three fiction selections and four prose passages with a deletion every eighth word. This deletion procedure was then modified when it was determined that certain blanks evoked ambiguous responses. These blanks were replaced in such a manner that only unequivocal words were deleted; the criterion for unequivocality was not stated. The tests were administered to children in grades three, four, five, six, and seven, using alternate forms of the same test. When the logarithms of estimated redundancies were plotted against each other, the relationships were found to be linear. It was concluded that for "skillful" readers the passages were "virtually completely redundant". The practice of selecting cloze deletions on other than a predetermined mechanical formula seems questionable since a random sample of the text involved is not obtained (see Table IV).

Bormuth (1967a) demonstrated a relationship between cloze and multiple-choice test scores. With an adequate range of subjects and carefully validated multiple-choice test items, Bormuth maintained that a 43 percent cloze score was equivalent to a 75 percent multiple-choice test score when corrections were made for guessing. In like manner, a 50 percent cloze score was shown equivalent to a 90 percent multiple-choice test score (see Table IV).

Based on Strickland's (1962) work on the patterns of oral language, Ruddell (1965) wrote six passages--three from high frequency syntactical patterns and three from low frequency syntactical patterns. Each passage utilized these specific patterns in the same proportional frequency with which they occurred in the oral language of fourth grade children. High frequency passages showed significantly higher cloze scores than low frequency passages. It was also concluded that cloze scores related to Stanford Reading Test scores, educational level of the parents, I.Q., and the chronological age of the subjects in the experiment (see Table IV).

Bormuth (1962) has compared the scores of 150 subjects on a specially-made comprehension test. One of the unique features of this test is that only words from the passage itself, or words that are known by 80 percent of the fourth graders in a similar population, are used in the test items. Three cloze passages from literature, three from social studies, and three from science--one at each grade level--are used with a population of fourth, fifth, and sixth grade subjects. Correlations between the comprehension tests and the cloze scores were found to be statistically significant (.946). Using a similar

Table III Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Deutsch 1964 Grade 1 N = 127 Grade 5 N = 127	1st Grade--verbatim samples of first grade teachers classroom speech, 18 sentences in length, final word deleted in each sentence. Sentences read to the child orally. 5th Grade--100 word verbatim speech sample from fifth grade teachers, 20% D.R. and two 50 word paragraphs from fifth grade children's speech. Group 1 (62 subjects) listened to deleted form, in its entirety once, and was allowed to write cloze word he thought deleted. Group 2 used conventional cloze reading procedure (pre-cloze). Scoring. EW, Contextually correct, and Grammatically correct.	Lorge-Thorndike I.Q., level 1 and 3 Non-verbal form A	Subjects: New York City Public Schools. Three-way analysis of variance. Dimensions: social status, race, and sex.	Cloze split halves reliability .33-.84. Cloze scores related to I.Q. ($\ll$.01). Main effect significance found favoring Hi SES and girls (most cloze measures). Only one instance of race main effect on cloze measures. Interactions (Hi SES & girls) found significant on many scores.

Table IV. Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
McLeod 1966 Grade 3-7 N = 100	Three children's fiction books, and four prose passages at the elementary level, every 8th word deleted; on the four passages (I), deletions which evoked ambig ous responses when tested on group of teachers; were replaced by ones that did not (II).	None	Between-subjects design. 2-way analysis of variance.	Group one scores on four passages higher than Group two score on three books. Means rank ordered by grade levels 3 through 7. Suggests that estimated redundancy of passages rated by two groups linear.
Bormuth (1967a) Grades 4-5 N = 100	50 deletions from each of nine 250 word passages, every 5th word deleted.	Nine 31-item multiple choice tests, one from each cloze passage. Items validated by 2 "experts". Items rejected if negative correlations between items score and total score found on 73 subjects.	Within-subjects: (order) cloze then M.C. after reading deleted passage. Tests for order effects insignificant.	Cloze and M.C. shown to be linear by observation. Regression equation shows cloze scores of 43% when corrected for guessing, related to M.C. score of 75%. Cloze of 50% comparable to M.C. of 90%. M.C. standard error of regression 6 points.
Ruddell 1965 4th grade N = 131	Six passages were specially constructed using same proportional word frequencies as oral language. Three high frequency and three low frequency patterns were equated (Dale-Chall) and deleted (D.R. 1:5). Passage length 254 words.	Otis Quick Scoring Mental Ability Test, New Edition, Beta Form EM. Stanford Achievement Test Intermediate 1.	Analysis of Variance w/repeated measures.	Split halves (+S.B.) reliability .85-.90 on cloze tests. Cloze and Stanford Reading Test scores related (r = .61-.72). Differences were found between high & low frequency patterns (<.01) on comprehension scores. Other variables significantly related to reading scores include education of parents, I.Q. and chronological age.

design, Bormuth (1964a) suggested that differences in cloze test forms over the same passage decreased as test length increased. A 1:5 deletion rate was used on each of the 250 word cloze passages (see Table V).

In three studies using the same type of passages (Bormuth, 1964c, 1965a, 1965b), cloze test scores were compared with Stanford Reading Tests and California Achievement Tests. Curvilinearity, optimum relations between number of subjects and number of items, and a variety of scoring patterns were investigated. A 1:5 deletion ratio was used in all experiments. Five different cloze tests were made over the same passage. It was found that no differences between test forms reached significance. At the word level and independent clause level, an analysis of regression showed relationships to be curvilinear, especially at the extreme ends of the distribution. Correlations between linguistic variables and comprehension difficulty were found to be significant. In the second investigation, the optimum relation of number of subjects and number of items was suggested. It was concluded that a combination of number of test blanks and number of subjects yielding a standard error of about 3.0 would yield stable scores. It was determined that a ratio of approximately 150 subjects to 50 cloze items would be appropriate, while it was suggested that an experimenter should use more subjects and/or items in actual practice, to allow for some margin of error. In the third investigation, Bormuth suggested that an exact-word scoring criterion accounted for 95 percent of the variance in cloze scores. While grammatically correct and synonymous responses correlated very highly with exact word scores and accounted for an increase in mean scores, little if any additional discrimination power was gained by using these relatively inefficient scoring criteria. It was concluded that the correlations increased as a function of the similarity of the meanings of the responses to deleted words (see Table VI).

In a recent investigation, Bormuth (1968) compared oral reading test scores and cloze scores to California Reading Achievement Test scores for fourth, fifth, and sixth graders. Supporting the findings of previous investigations (Bormuth, 1967a), cloze scores of 44 percent were found to relate to reading achievement test scores of 75 percent, while cloze scores of 57 percent related to 95 percent on the reading achievement tests. Cloze scores of 33 percent and 54 percent were found comparable to a 75 percent and 95 percent score on oral reading (word recognition) tests. Due to much greater variance in oral reading (word recognition) scores, it was suggested that word recognition and comprehension criterion scores are not comparable. Large differences between the relationship of cloze and oral reading scores were found (see Table VII).

Schneyer (1965) compared a cloze test in which the deletion rate was 1:10 to scores on the California Test of Mental Maturity and the Gates Reading Survey. It was found that I.Q. was related significantly to cloze scores. No increase in scores on the Gates Reading Survey resulted from practice with a series of 200-word cloze passages when

Table V Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions scores (main effects, interactions)
Bormuth 1962 4th, 5th & 6th grade N = 150	Nine passages, 3 literature, 3 social studies, 3 science; three levels of difficulty, 4.5, 5.5, 6.5, (Dale-Chall); length was 270 to 299 words. Deletion rate 1:5.	Comprehension test 4 option, multiple-choice, 31 items/passage covering material given in cloze tests. Only words from passages appeared in test items or were known by 80% of 4th graders.	Subjects from small midwestern towns. Correlations tested for significance.	Positive correlation between comprehension and cloze tests; significant. Item difficulty range .00-1.00. Validity of cloze as a readability measure seems likely.
Bormuth (1964a) N = 139 Grade 4-8	Five 50-item tests made from same passages such that all words were deleted in one of the 5 forms. First sentence not deleted; deletion rate 1:5. Five subject areas tested:literature, history, geography, biological science, physical science. 20 passages.	Stanford Reading Test	Groups matched by Stanford Reading test means. Every subject in each group took one form of each cloze test. Correlations tested for significance.	Differences in difficulty among test forms made from the same passage tend to diminish as more items are included. Standard test form error designed to aid in determining between-forms differences.

Table VI Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Bormuth (1964c) N = 650 Grades 4-8	20 passages, 275-300 words each. Deletion rate 1:5. Five forms for each passage made by starting deletions with first word (A) through fifth word (E). Scoring: EW. Separate scores calculated for three linguistic variables: word, phrase, and passage.	Stanford Reading Achievement Test Form J	All subjects are from small California school district. Subjects stratified into 5 reading ability levels on Stanford scores. 3X5 analysis of variance used to determine relation of linguistic variables and reading ability levels.	Differences between forms (A-E) not significant. Word level and independent clause level of analysis: regressions found curvilinear, most measures significant especially at extreme ends of distribution. No F-ratio reached significance at passage level but curvilinearity is suggested. Multiple correlations found between linguistic variables & comprehension difficulty .51-.68
Bormuth 1965 (a) N = 695 Grades 4-8	20 passages, deletion rate 1:5. Scoring: EW. Content of passages included: literature, geography, history, biological science, physical science. Difficulty level grades 4-8 (Dale-Chall). 250 words in length. Tests were divided about every 5th deletion to give 10 test lengths.	California Achievement Test	Small Calif. school district 86% caucasion, 10% negro, 4% Mexican. Two cloze tests given each day for 10 days.	Average standard error calculated for number items and number of subjects. Optimum relation between number of subjects and items suggested. Standard deviation of cloze tests ranged from 4.5 to 9.1. Aludes to assumptions underlying Lord's formulas but does not discuss.
Bormuth 1965 (b) N = 50 Grade 5 & 6	20 passages, 52 cloze deletions, deletion rate 1:5. Scoring criteria: Exact word (EGC), synonyms (SGC), grammaticality (SGI), unrelated (UGI), and unclassifiable (UCR). Total of 1040 items considered as single test.	Stanford Achievement Test: Reading	Multiple regression analysis using reading achievement scores as criterion.	EGC scores accounted for 95% of variance. Grammaticality correlated possitively with comprehension scores. Correlations increased as a function of the similarity of the meanings of the responses to the deleted words.

Table VII Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Bormuth 1967 (c) Grades 4, 5, 6. N = 120	Four 13 paragraph graded sequences (pre-primer through adult) were deleted 1:5; two deletions forms scoring EW.	California Reading Achievement Test grades 4-6. Two of the four paragraphs were taken as cloze tests, two as oral reading exercises (scored as word recognition exercises).	All subjects from single school. Correlations obtained and tested for significance.	Reading test score ranged from 2.0 to 11.3. 1. Cloze scores of 44% and 57% found comparable to 75% and 95% on comprehension scores (C.R.T.). 2. Cloze scores of 33% and 54% comparable to 95% on oral reading (word recognition) scores. 3. Large differences between the relationship of cloze and C.R.T. and cloze and oral reading scores. 4. Suggested that word recognition and comprehension criterion scores not comparable.
Schneyer 1965 Grade 6 N = 66	200 word passages from basal reader. Deletion rate for treatment A, 1:10. Deletion for treatment B nouns and verbs only (alternately on average 1:10).	California Test of Mental Maturity. Gates Reading Survey Forms I & II.	Comprehension scores covaried with initial scores.	Gain scores not significant. IQ related to 10th word deletion scores ($<.01$) and noun-verb deletions (.02). All cloze scores and standardized reading measures related ($<.01$). Two cloze scores related ($<.01$). Suggests that noun-verb cloze scores may be less related to IQ than every nth cloze scores.

tests were corrected and immediately returned to the subjects. Two types of deletions were used: every 10th word and noun-verb-only deletions (see Table VII).

Louthan (1965) found no differences on comprehension test questions between cloze test scores of 7th grade students when deletions were: (a) every 10th word, (b) only nouns, (c) only verbs, or (d) only adjectives. Louthan argued that cloze tests have limited value as a teaching instrument. A control group reading undeleted material, received superior scores on the comprehension test to groups in which every 10th word or only nouns, verbs, adjectives, or other modifiers, were deleted. However, when only prepositions and conjunctions were deleted, or when only pronouns or noun determiners were deleted, cloze scores surpassed the groups mentioned above. These data suggest that the so called "content word" deletions (class 1, 2, 3, 4) require different skills than "function words" (class 5, 6, 7) (see Table VIII).

Bloomer (1965) compared posttest scores of groups who received: (a) pretest and undeleted material, (b) only deleted material, (c) a pretest and deleted material, and (d) only undeleted material. That group which did not receive a pretest and did use deleted material, performed significantly better than any other group. One somewhat questionable conclusion of this study was that anxiety caused by the pretest accounted for the results shown (see Table VIII).

In a subsequent study of specially-written multiple-choice test scores given after 24 cloze exercises, Bloomer (1966) concluded that sex differences disappeared with increasing age and that cloze tests made from very easy materials were less "motivating" than materials more closely associated with grade level of the subjects. In a longitudinal study of ninth graders, Bloomer argues that cloze procedure is proven relatively ineffective in producing an increase in reading comprehension scores (see Table VIII).

Section B--adults

The studies in this section are organized into five groups: (a) the nature of contextual constraint, (b) the methodology of cloze procedure, (c) literary style and the cloze procedure, (d) the cloze procedure and information gain, (e) the cloze procedure and oral speech. The studies within each section are organized chronologically.

The nature of contextual constraint. A study by Aborn, Rubenstein, and Sterling (1959) investigated the constraint upon words attributable to the number of words between deletions, distribution of deletions, and structure of context. One word was omitted from each sentence in a way that yielded three treatments of sentence length, four treatments of position of omission, and six treatments of word class. The results noted in Table IX indicate that bilaterally distributed context of over five and under ten words yields cloze scores of maximum

Table VIII Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Louthan 1965 N = 236 Grade 7	24 passages 500-600 words each one undeleted form, 7 cloze forms: by form class (10% deletions): 1) every 10th word 2) nouns 3) verbs 4) modifiers, adjectives, adverbs 5) prepositions and conjunctions 6) noun determiners 7) pronouns Control) undeleted	Comprehension test: 12 questions for each passage: multiple choice.	Correlations tested for significance.	No differences in comprehension test scores between class 1, 2, 3, 4. Column A significantly higher (<.01) than column B: A: control — B: 1, 2, 3, 4, 6 A: 5 — B: 1, 2, 3, 4 A: 6 — B: 1, 2, 3, 4, 5, control A: 7 — B: 1, 2, 3, 4, 6
Bloomer 1965 N = 146 Grade 8	Deletion rate 1:10 on five most difficult passages of McCall-Crabbs Reading Test.	Identical pre- and post-test. Nature of pre-test not given. Completion times recorded for cloze test, pre- and post tests.	Subjects drawn from central N.Y. Four groups: A. Pre-test — undeleted material B. none — undeleted material C. none — deleted material D. pre-test — deleted material Correlations tested for significance.	1. No differences between (A & D) or (B & C). 2. C superior to A or B or D. 3. A and D combined superior to B & C combined. 4. D and C took longer than A and B. Argues that pretests should not be used.
Bloomer 1966 N = 976 5th, 7th, 9th, 11th grades	Twenty-four 600 word passages, fifth grade difficulty.	Special comprehension test: 12 question multiple-choice admin. after each cloze passage.	All subjects from a single school district, correlations tested for significance.	11th graders-lower scores than 9th graders. Sex differences disappear as age increases.

discrimination power.

Using thirty-six, 150-word passages ranging in difficulty from first grade level through technical prose, Miller and Coleman (1966) investigated three deletion techniques: (a) mechanical deletion rate of 1:5 (yielded the highest standard deviation), (b) the deletion of only one word in each passage (yielded the highest means), and (c) a unilateral constraint system in which each succeeding word was guessed and then revealed to the subject. Method c showed the lowest scores. Method c, however, revealed that sequential constraint was strong within sentences while little constraint was shown to cross sentence boundaries. Insignificant increases in performance curves were evident when the context surrounding deletions exceeded 20 words. The authors concluded that cloze scores reliably measured readability ranging from first grade to adult levels (see Table IX).

The methodology of cloze procedure. In the first comprehensive statement of the cloze procedure, Taylor (1953a) investigated methodological questions relating to the potential reliability and validity of the cloze tests. This research, discussed in the introduction to this paper, will only be summarized here. Using two 175-word fiction passages, every nth (1:10) and random 10 percent mechanical deletion systems were considered, scoring both exact word and synonym replacement separately. It was concluded that; (a) cloze scores ranked passages in the same way as readability formulas, (b) the two deletion procedures tested, yielded reliable discrimination power between subjects, (c) a minimum of 35 blanks with a deletion rate of 1:10 discriminated "better" than other less frequent deletions or fewer blanks per passage, (d) scoring systems other than the exact word scoring system yielded virtually identical scores. In a second part of the same investigation, Taylor (1953b) validated the findings of the first study and suggested that reading abilities of individuals might be assessed by utilizing the cloze method, since significant between-subjects F-ratios held up across passages (see Table X).

Fletcher (1959) investigated the relationship between three 250-word cloze passages (deletion rate 1:5) using an exact word scoring criterion. He found that: (a) there was a substantial positive relationship between a subject's ability to use context clues and ability to comprehend rapidly, (b) ability to use context clues was significantly related to "general intelligence", or more specifically to the verbal factors of "general intelligence", (c) appropriate cloze procedures can be developed as the basis of a test to measure a subject's ability to use context. It was argued that this measure was both valid and reliable. In addition, the instrument would be easy to prepare (see Table XI).

Constraints within complete prose paragraphs were investigated by MacGinitie (1961) by comparing restorations of omitted words when context at different distances from these words was deleted. Fifteen

Table IX Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Aborn, et. al. 1959 N = 24 freshmen	Words from 1,380 six, eleven, and twenty-five word sentences from popular magazines were deleted from beginning, early medial, late medial, and final positions. Scoring--E.W., misspellings correct.	None	Selected from the top 10% on reading and English test and top 25% on S.C.A.T. Tested every day for 4 weeks, 1.5 hrs/day.	1. Length and distribution of context are independent sources of constraint. 2. Predictability of words inversely related to size of class. 3. Context over ten or under five words produces maximum constraint. 4. Bilateral context distribution is optimum. 5. When frequency or word class is uncontrolled predictability similar for all deletion positions except final which is lower.
Miller & Coleman 1966 N = 479 college students	Thirty-six 150 word McCall-Crabbs test passages from 1st grade through difficult technical prose. Three deletion forms (a) D.R. 1:5 (b) only one word/passage, (c) unilateral constraint only (each succeeding word guessed then revealed). Scoring--E.W. for (b) and (c); (a) E.W. weighted 3, syn. as 2, form class as 1.	None	Treatments: (a) 20 S's allowed 6 min/passage. Order of presentation randomized. (b) 450 S's three for each version of each passage. (c) 9 S's, new word exposed every 7 seconds.	Synonym and form class scores correlated .99 with E.W. Mean S.D. a) 54.6% 14.5 b) 63.8% 11.0 c) 33.7% 7.6 Correlations: a with b----------.95 a with c----------.87 b with c----------.87 (C) shows sequential constraint within sentences. (Tau for 8 positions .79, .01) Little constraint across sentence boundries. From a given position no increase in performance curve after 20 words.

Table X Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Taylor 1953 I N = 24 adults	Three 175 word (approx.)fiction passages, Flesch reading ease 47, 68, 89; Dale-Chall 6.4, 7.1, 9.2. Each passage deleted 1:10 and random 10%. Scoring E.W. Separate synonym score counted 1/2 E.W. score.	None	Order of presentation controlled. Analysis of variance.	1. Cloze scores ranked passages same as Flesch and Dale-Chall formulas in every instance. 2. Significant F ($<.01$) between difficulty levels. 3. All deletion procedures yielded reliable discrimination power between subjects. 4. 35 blanks, D.R. 1:10 discriminated "better" than other less frequent deletions or fewer blanks/passage. 5. Synonym scoring yielded "virtually identical differentiation." 6. Presentation order had no effect on scores.
Taylor 1953 II N = 72 adults	All passages above (Taylor, 1953 I) and seven additional 175 word fiction and non-fiction selections used. Each passage deleted 1:5, 1:7, 1:10, 10%. Flesch and Dale-Chall rankings differed on additional passages.	None	(see I above)	1. Findings of #1 validated: Some passages ranked in same order. 2. Cloze ranked passages more accurately than Flesch or Dale-Chall. 3. Cloze rankings consistant across subjects. 4. Deletion method and rate findings inconclusive. 5. Reading abilities of individuals might be assessed since significant between subjects F scores held up across passages.

Table XI Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Fletcher 1959 N = 174 Freshmen	Three 250 word, D.R. 1:5. Scoring--E.W. Fiction; difficulty for each passage (Lorge Formula) was 5.63, 6.99, and 6.94.	(a) Dvorak-Van Wagenen Rate of Comprehension and (b) Cooperative, Speed of Comprehension Test, (c) Cooperative Level of Comprehension Test, (d) Cooperative Vocabulary Test, (e) American Council on Education Examination of College Freshmen.	Correlations calculated	Reported Correlations: 1. Cloze and rate (speed) of comprehension measures scores (a) .59 and (b) .57. 2. Cloze and level of comprehension score (c) .54. 3. Cloze and vocabulary (d) .63. 4. Cloze and general verbal ability (e) .72.
MacGinitie 1961 N = 600 College Students	Two 144 word descriptive narratives with Lorge readabilities of 4.9 (A) and 7.0 (B). Fifteen different omission sets in nine patterns used with each passage: 1:3, 1:6, 1:12, 1:24, starting on first word; above patterns are repeated starting first minus one word, 2:6, 2:12, 2:24 starting on first word; 2:6, 2:12, starting on first word minus three; 2:13 starting on first word minus two; 4.24 starting on the first word. Scoring--E.W.	Vocabulary test--G.T. Form 1 & 2.	Pretest (Form 1 & 2) followed by cloze passages. Mean of Subjects on American Council Psychological Exam in top 25%	1. No differences were significant between cloze scores from tests using 1:6, 1:12, 1:24 deletion rates. 2. A deletion rate of 1:3 made restoration more difficult. 3. Omitting words in pairs made cloze test scores lower under all conditions.

different omission sets, in nine patterns, were used with two 144-word descriptive narratives of sixth grade difficulty level with 600 college students. It was concluded that words were equally restorable every 24th, 12th, or 6th word, but that third-word omissions made restoration significantly more difficult. It was further concluded that the omission of two adjacent words increased difficulty but that the two words were equally restorable. The influence of context upon word choice, it was suggested, decreases rapidly with the distance of the context from the deletion up to about five words. Distance beyond five words, it was argued, has less effect on cloze scores (see Table XI).

Weaver and Kingston (1963) conducted a factor analyses of cloze procedures and other measures of reading ability. Contrary to previous investigations, it was concluded that the relationship of cloze tests to standard tests used in this study was only moderate. By means of a rotated factor analysis, three factors were extracted from the 18 tests used which were identified as verbal comprehension, redundancy utilization, and (more tentatively) rote memory. The authors point out that there is much specific variance connected with cloze test scores that, in their investigation, was not accounted for by standardized tests of reading comprehension, listening, and language symbol manipulation (see Table XII).

Fillenbaum (1963) obtained cloze scores from transcripts of speech from Thematic Apperception Test responses using varying deletion rates (1:2, 1:3, 1:4, 1:5, 1:6); 200 deletions were made. Cloze responses were considered as correct if they were verbatim or from the same form class as verbatim responses. The relation between these two measures was examined. The main findings of this investigation were that cloze scores increased moderately with decreasing density of deletions. It was argued that the determinants of form class and verbatim items in cloze completions are different since form class predictability is more dependent upon the immediate grammatical environment whereas verbatim predictability depends upon both this factor and remote topical content or semantic features of discourse. It was also suggested that these determinants vary with the particular grammatical class under consideration (see Table XIII).

An investigation by Hafner (1963), of different methods of scoring cloze tests, considered five different measures: (a) connective words only, (b) content words, (c) connective/content quotient, (d) grammatically correct but lexically incorrect responses, and (e) incongruent scores (responses which are contrary to clearly stated information in the context or language pattern of the sentence in question). These measures were compared with three standardized measures, two experimental tests, and course grades in reading methods for college seniors. The major findings of this investigation were that cloze scores correlated positively and significantly with all the standardized measures. The incongruency measure correlated negatively and significantly with

Table XII Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Weaver & Kingston 1962 N = 160 College Juniors	Eight 40 item cloze tests included: Printed passages: Materials — Deletions a) Essay — Mechanical b) Essay — Form class c) Speech — Mechanical d) Speech — Form class Spoken passages: Materials — Deletions e) Essay — Mechanical f) Speech — Form class g) Essay — Form class h) Speech — Mechanical	1. Davis Reading Test Modern Language Aptitude Tests: 2. Number Learning 3. Phonetic Script 4. Spelling Clues 5. Words in Sentences 6. Paired Associates 7. STEP Listening Test Ohio State Psychological Exam 8. Vocabulary 9. Word Relations 10. Reading Comprehension	Within subjects design, no control for order effects mentioned. 30 minutes allowed for each cloze test (written or oral). Listening cloze tests; subjects heard entire passage first then only context preceeding deletion. Correlations calculated and rotated factor analysis for I Reading Comprehension II Cloze Factor III Memory and analogy	Relationships with cloze scores on spoken passages accounted for 70% of significant relationships over .50. No correlation with cloze scores exceeded .61. Relationships cloze scores and Ohio State Psychological Examinations accounted for 70% of the correlations over .50. The most heavily loaded factor is reading comprehension (24.83) followed by (3.93) and memory and analogy Cloze tests were found only moderately related to verbal comprehension. Suggests that an aptitude distinct from verbal comprehension underlies cloze scores on printed materials. Much specific variance connected with cloze scores unaccounted for.

Table XIII Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Fillenbaum, et al. 1963 N = ? undergrad	Transcript of speech from T.A.T. responses. Deletion rates 1:2, 1:3, 1:4, 1:5, 1:6. Two hundred deletions made at each rate. Scoring--E.W.	None	Cloze scores considered by grammatical class. Subjects tested in groups of 5 to 20. Order of test presentation randomized.	Proportion of verbatim completions: Content words / Function words / Total 1:2 .115 .483 .323 1:3 .245 .581 .434 1:4 .297 .606 .464 1:5 .319 .646 .514 1:6 .321 .704 .540
Hafner 1963 N = 32 college seniors	One 250-word article on educational television. D.R. 1:5. Scoring: A. Cloze E.W. B. Incorrect but correct form class (G.C.I.A.) C. Connectives correct D. Connective/Content Quotient E. Contrary to contextual information or language pattern (Incongruent) F. Cloze Speed	1. Michigan Vocabulary Profile 2. Otis Quick Scoring Mental Ability Test, Gamma Fm. 3. Weschesler-Bellevie Intelligence Scale, Information (sub. T) 4. Hafner Intelligence Test (experimental) 5. Hafner General Information Test (experimental) 6. Course Grade (Reading Methods)	Correlations tested for significance.	Cloze scores related to intelligence, vocabulary, information and course marks. Cloze speed related to information, achievement, vocabulary and intelligence. Cloze variables did not enter a Multiple R prediction of college grade point average. Relations (<.01) (*<.05). A B C D E F 1 .56 N/S N/S .-42 .-50 .48 2 .73 .68 .69 N/S .-48 .64 3 .56 N/S N/S N/S N/S N/S 4 .46 N/S N/S N/S N/S N/S 5 N/S *42 N/S N/S .-47 *41 6 .65 .47 N/S N/S .-54 *41

vocabulary intelligence measures and course grades. It is hypothesized that the ability to complete blanks where function words have been deleted was an indication of the ability of the subject to conceptualize interrelationships among ideas. Personality variables were also considered in this investigation (see Table XIII).

In a study by Musgrave (1963) a novel story was presented to 200 subjects. In one form, the story was presented with information on plot and character; in another form the story was presented without the information (which was contained in a paragraph preceding the story itself). Cloze scores were derived from both exact word scoring and "commonality scoring" (defined by the author as words which are exact duplicates of the most popular responses made by the group of subjects under consideration). It was found that exact word and commonality scores correlated highly with each other and that there were no significant differences in those passages preceded by "who" explanations or "what" explanations (see Table XIV). As was suggested by the author, however, these findings may have been confounded by a crucial variable, the topic of the story itself. If, as Coleman suggests, contextual constraint is more powerful in determining cloze completions than is prior information, a successful demonstration of the effect of these context factors would not be possible.

Luke (1964) investigated deletions (10 percent) in 150-word fiction passages. These deletions were of four types: nouns only, verbs only, adjectives only, and combinations of the three. In this investigation the mean scores for verb deletions were higher than combination scores--which were in turn higher than either nouns or adjective means. The seeming contradiction of Louthan's (1965) data may lie in passage differences or in the age of subjects (7th grade vs. adult) (see Table XIV).

In another investigation of different types of deletion methods, Greene (1965), using a 600-word passage, deleted only content words in a mechanical 1:12 manner and in an 8 percent mutilation method. The content word tests showed higher scores than those tests that were mechanically deleted. The author points out that while this modified cloze procedure did produce higher means, test construction time is greatly increased and there is a loss of objectivity in item construction. It may also be pointed out that rational deletion ceases to measure passage difficulty since the deletions are no longer representative of the population of possible deletions within the sample (see Table XV). It may be noted that an investigation by McLeod (1965) has resulted in the publication of a reading test (GAP Reading Comprehension Test, 1965) based on a cloze type procedure using rational deletion of certain words. The instrument is designed for use with children of upper elementary school age.

Literary style and cloze scores. Using Miller and Selfridge's (1950) 50-word passages (rearranged in eight orders from highest to

Table XIV Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Musgrave 1963 N = 200 undergrad	Three 200 word newspaper stories. D.R. 10% random. Scoring--E.W. and "commonality".	None	Treatments I. Only simple instructions & deleted passage. II. Same as I but first para not deleted (told "who" and "what"). III. Same as II but only "who" portion of lst para included. IV. Same as II but only "what" portion of lst para included. Analysis of variance.	Mean for "commonality" scores ranged from 7.92 - 8.76 out of 20 while E.W. means were 6.04-6.54. Treatment II and Treatment III means somewhat higher than IV or I for both "commonality" and E.W. scoring criteria. No significant main effects or interactions.
Luke 1964 N = 24 female undergrad	10% of four 150 word fiction passages were deleted four ways: a) nouns only b) verbs only c) adjectives only d) combination (5 each) of the 3 form classes above. Scoring--E.W.	None	Each of the four deletion types were administered so that no subject received two cloze tests over the same passage. Only percentages reported.	Higher cloze scores reported for verbs than nouns or adjectives (M 6.12). Combination scores (M 5.29) were followed by nouns (M 5.00) and adjectives (M 4.20).

Table XV Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Greene 1965 N=128 Upper classmen & grad. students	One 600 word post cloze passage. Form B-C deleted 1:12; Form A-D 8% deletion, content words only.	None	65 S's took B-C, 63 S's A-D	Split half reliability of B-C (.523) and A-D (.757) significant (<.05). Item discrimination (r over .40) greater for A-D (64%) than B-C (55%). Modified cloze (A-D) produced higher percentage of "effective" items. 13% of items in B-C extremely difficult.

lowest statistical approximation to English), Salzinger, Portnoy, and Feldman (1962) found that the proportion of words guessed correctly in any grammatical category increased as the approximation to English word order increased. In this study, which is highly related to communication theory, Salzinger at al. concluded that the assumption of equal intervals between successive orders of approximation is untenable when using cloze scores as a criterion. The results further indicated that the relationship between memory and order of approximation can be explained in part by the syntactical structures themselves and in part by meaning called for by the context (see Table XVI).

Coleman has investigated two aspects of the structural characteristics of passages and, with the use of the cloze technique, has evaluated their relative readability. In one study (Coleman, 1962), three 232-word passages from adult non-fiction were rewritten so that sentence length varied from 15 words to about 38 words. Subjects were matched in groups of three and each was instructed to read one of the passages. It was found that, while increased sentence length accompanied lower cloze scores for that sentence, differences were small. These results indicate, according to Coleman, that the indiscriminate break up of all long sentences into short ones will improve the passage either slightly or not at all (see Table XVI). In a second investigation (Coleman and Blumenfeld, 1963), two passages from a psychology text and ten selections from technical writings were presented to subjects in both nominalized and active verb forms. Cloze tests were composed using a 1:5 deletion rate and each form was given to half of the subjects. Function words and content word scores were analyzed separately. Differences in cloze scores favored the active verb form. Content word scores for the active verb form were superior to content word scores for the nominalized sentences. It was concluded that the word patterns of the material written in active verb form were more predictable, thus allowing the individual, who is required to read and gain information from a passage, to know relatively more about the material after the first reading. It was suggested that these results have implications for preparing teaching material (see Table XVI).

Without specifically defining the characteristics of the passages used, Bormuth (1965c) has compared the cloze scores on two author's material with the scores that subjects received on tests of ability to detect literary style. Correlations with the style test were significant. However, no differences between correlations for the two passages approached significance. Bormuth seriously questions Weaver and Kingston's (1963) conclusions and suggests that because they, like Rankin, have used a selective deletion procedure, factor analyses fail to load heavily or uniformly on the same factor (see Table XVII).

Darnell (1963) compared the cloze scores of seven binary transpositions of a 240-word passage. It was logically determined that the seven degrees of disorganization resulted in a loss of clarity. The findings showed that cloze scores ranked the passages in a manner that

Table XVI Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Salzinger et. al. 1962 N=93 undergrad	Miller and Selfridge's (1950) nine 50-word passages in each of eight orders of statistical approximation to English were deleted 1:5. Scoring--E.W. and form class.	None	Between subjects design.	1. Subjects guessed a greater proportion of words the higher the order of approximation to English. 2. Proportion of words guessed in the correct grammatical category increased most at the end of orders (0-3) furthest from English.
Coleman 1962 N=90 Undergrad	Three 232 word passages from adult non-fiction. Each re-written to give three sentence lengths: Group A 15.4 words B 23.2 words C 38.7 words A cloze test was made over original passage.	None	Each subject read one passage in each group, but each read different group-passage combinations.	Mean scores A - 22.4 B - 21.3 C - 20.9 Linearity significant (<.05) The degree by which the cloze scores of A & B exceeded C correlated with length of C. r.383 (<.05) Argues that indiscriminate shortening of sentences may not necessarily improve the passage.
Coleman & Blumen-feld 1963 N=100 freshmen	Two paragraphs from a psychology text and ten sentences selected from technical writings, were written in nominalized and active-verb form and divided evenly among the subjects. The 367 words were deleted (D.R. 1:5) in ten forms so that each word was scored for nominalized and active verb scores. Scoring--E.W. and inflected forms correct.	None	Each subject received half nominalized and half active-verb sentences. Analyzed for function word and content word scores.	Active-verb sentence cloze scores superior to nominalized cloze scores (<.01). Content word scores for active-verb superior to those for nominalized sentences (<.01).

Table XVII Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Bormuth, et. al. 1965c N=150 female college students	Two 250 word selections from Trollope and two from Steinbeck. D. R. 1:5, scoring--E.W.	Test of ability to detect literary style (unidentified).	First cloze test followed by reading and discussion of passage followed by second cloze test and test of ability to detect literary style.	1. Cloze test reliability (split half and S.B.) .76 - .94. 2. Correlation with style test .418 - .660 (<.01). 3. No differences between correlations for Trollope and Steinbeck approached significance. 4. Study of undeleted passages does not seem to effect scores on measures to detect literary style. 5. Pre-cloze and post-cloze seem to be equally valid.
Darnell 1963 N=140 Undergrad	One 240 word control passage and one similar passage in which sentences are rearranged in seven degrees disorder, 1 ("right") to 7 (most disordered). Number of binary transpositions for each form: 1 0 2 7 3 13 4 20 5 27 6 35 7 43	None	7 experimental groups (20 S's ea.) took the control cloze test and a cloze test over one of the seven transposed passages.	Seven forms were significantly different (<.05). Group 1 2 3 4 5 6 7 Control Mean 23 22 22 25 22 23 21 Exper. Mean 21 19 17 18 18 18 1[illegible] Cloze % Exper. 45 40 37 39 38 38 32 1. Disorder can affect comprehension adversely. 2. Amount of loss of clarity becomes greater as the degree of disorganization increases.

reflected the amount of disorganization caused by the transpositions. Darnell concluded that disorder can affect comprehension adversely and that the amount of clarity lost increases as disorganization increases. It is suggested that one strength of the cloze procedure is that it is sensitive to the organization variable and can be used in further studies of the effect of word and sentence order on comprehension (see Table XVII).

Cloze scores as a measure of knowledge or information gain. Taylor (1956a) gave a group of Air Force trainees a cloze test, then allowed them to read the passage in the undeleted form. This was followed by another form of the same cloze test. While the pretest and posttest scores correlated highly, the gain was significant. In the same study, Taylor investigated a methodological question and, by comparing function word scores and content word scores (passages deleted 1:5), he found that function word deletions produced higher means but showed smaller gain scores than content words. Correlations between cloze and comprehension test gain scores were highly significant for contentives. However, because of the high correlations between both content and function word scores, Taylor suggests that there is little advantage in preclassifying words and limiting deletions to them, and no advantage in putting one's self to the trouble of judging and scoring synonyms. In this same investigation, Taylor concluded that about 50 blanks were sufficient to allow the chances of mechanically selecting easy and hard words to cancel out and yield a stable score of the difficulty of the passage or the performance of an individual despite specific words of extreme ease or difficulty (see Table XVIII).

The findings of Bloomer (1962) suggested that when a standardized instrument was used to measure possible gains from cloze test completion that gains, if any, were not significant. Bloomer reported that one control group was completely eliminated because the volunteer subjects failed to complete the experiment. This suggests that the findings from the other groups involved may be affected by the manner in which subjects were selected (see Table XIX).

Coleman (1966) used a pre-cloze test and a post-cloze test as measures of information gain. While certain gains were reported, none of statistical significance was found. Coleman speculated that this was due to the nature of the cloze as a pretest measure by suggesting that the deleted paragraph contains substantial information (see Table XIX).

The effectiveness of oral communication as measured by cloze scores. Dickens and Williams (1964) compared scores on the American College English Abilities Test with scores from an orally administered cloze test. It was indicated that cloze procedure was a useful technique for measuring the comprehension of oral messages. Correlations with the English Abilities Test were significant and it was suggested that the cloze test taken orally was both reliable and valid. Oral cloze

Table XVIII Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Taylor 1956 I N=48 Air Force Trainees	Eight 9 line samples of an Air Force supply manual were deleted (D.R. 1:5 implied).	Specially constructed comprehension tests with reliability "checks" (no data) and item analyses (no data). Pre & post test forms used. Air Forces Qualification Test also compared.	Subjects given one cloze form, then pretest one. One week later given undeleted article to read followed by the post test & another copy of the cloze form. Correlations.	1. Pretest scores correlated .70 and post test scores correlated .80 with each other. 2. Pretest-post test relation .83 (comprehension) and .88 (cloze). 3. Comprehension scores correlated .65 (before) and .70 (after) with A.F.Q.T. 4. Cloze scores correlated .73 and .74 with A.F.Q.T. 5. Mean gain 8.5 points for cloze and 4.8 for comprehension (<.001).
Taylor 1956 II N=104 Air Force Trainees	Similar to Taylor 1956 I but deleted only functors or contentives.	None	(see above)	1. Gain scores significant for both functors and contentives. 2. Functors produced higher means but smaller gain scores than contentives scores. 3. Contentives showed less stability than mechanical deletion procedures. 4. Correlations between cloze and comprehension tests gain scores was .80 for contentives only and mechanical deletions.

Table XIX Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Bloomer 1962 N=88 college students	500 word passages from each of seven levels of a commercial science and social studies series. D.R. 1:10	Diagnostic Survey Test Form A & B	Control group I used Gilbert's Power and Speed ...l Reading Control. Control Group II took pre & post tests (Diagnostic Survey) only. Experimental group took pre test, post tests and cloze exercises.	1. All but 2 subjects in control group I dropped out. 2. While gain scores for both control group II and the experimental group were significant no between group differences were significant. 3. Total and comprehension scores favored the experimental group but vocabulary measure favored control group II (non-significant).
Coleman 1966a N=20 Undergrad	Thirty-six 150 word McCall-Crabbs test passages from first grade through difficult prose. Scoring--E.W., D.R. 1:5.	None	Information Gain for each passage is defined as difference between pre-cloze and post-cloze scores. S's took cloze test then read undeleted passages then another form of cloze test.	Mean gain 1.85. No significance, it is contended is due to nature of cloze as a pre-test.

procedure may have some advantages over multiple-choice tests as a research technique for studying the comprehension of spoken messages. DeVito (1965) obtained cloze scores on samples of oral and written discourse by skillful communicators. It was suggested that the passages used did not differ in overall comprehensibility, in vocabulary load, in sentence structure, or density of ideas. No differences between groups was reported. It may be that the cloze procedure is not a sensitive device when vocabulary and other characteristics are relatively advanced and the respondents are comparatively less able (see Table XX).

SUMMARY AND CONCLUSIONS

THE SUMMARY--NATURE OF CLOZE

Cloze measurement appears applicable to many types of communication. It seems to discriminate among the "readability" levels of passages and among the reading comprehension levels of readers. The cloze technique provides a measure of the degree of correspondence between the language habits of the transmitter and those of the receiver.

Although the cloze procedure measures, in a sense, the degree to which the transmitter and the receiver share the same structural and lexical communication patterns, the "total stimulus complex" of each may be quite different at any specific point in the communication. The transmitter is combining words both on the basis of previous experience and on the basis of the goal toward which his writing is aimed. The receiver, on the other hand, may lack a foreknowledge of this goal, thus the nature of his receptions tends to be limited only by his own past experience.

It has been suggested (Taylor, 1956) that pairs of words that have a high or a low probability of occurring together, greatly influence cloze scores. The most frequent sequential associations would tend to limit the population of possible cloze responses. Thus, high probability messages would withstand mutilation better than low probability ones. While this suggests that the most readable message is, therefore, one that contains the highest possible number of high probability associations, it may be that under these conditions little or no new information is transferred and no novel phrase or concept would be presented. One may ask the question then: "Why should such material be read at all?"

Some degree of redundancy must exist, however, if the receiver is to understand the message being tansmitted. A complete lack of redundancy would tend to create total disorganization. It has been shown that highly disorganized messages have little or no communication value (Salzinger, 1962; Darnell, 1963).

The summary of studies of children. Studies of children using the

Table XX Investigations Involving the Cloze Technique

Author Date Subjects Age/ grade, number	Cloze Test Description type of material, difficulty level, deletion rate, word class deleted, scores.	Other Measures name of instrument, type of scores	Design subject selection, subject description, treatments, controls	Findings relationships, conclusions, scores (main effects, interactions)
Dickens & Williams 1964 N=253 undergrad	Two 528 word speeches from the STEP. Selection A-one word mutilated every 5 seconds. Selection B-one word mutilated as in section A and 5 second silence for responses at end of each sentence.	American College Testing Program, English Abilities Test.	Group 1 (N=127) heard cloze passage as deleted in its entirety, then on second hearing responded. Group II (N=126) heard mutilated tape followed by 32 mutiple choice questions taken in part from STEP Listening Test and in part from investigator's work. Correlations tested for significance.	1. Cloze scores correlated with English abilities scores .49 & .52 (<.01). 2. Two passages cloze scores related to each other .73 (<.01). 3. Multiple choice test scores correlated with English Ability scores .48 & .50 (<.01). 4. Two passages multiple choice scores related to each other .37 (<.01). 5. Cloze test split half reliability .70 & .80.
DeVito 1965 N=85 16-18 yr. olds	Four 300 word passages (2 oral, 2 written), D.R. 1:5. Scoring--E.W.	None	Groups of 16-18 Subjects took each passage.	Pairs of samples of oral and written discourse by skillful communicators do not differ significantly in overall comprehensibility. Cloze scores mean / S.D. Oral 19.84 / 3.48 Written 19.86 / 3.42

cloze technique began in the early 1960's and were first reported in 1962 (Bormuth). Sources of the passages selected for deletion included standardized tests, conversation of adults and children, fiction and non-fiction selections from authors other than the investigator, and paragraphs specially constructed by the investigator. Research involving the cloze procedure has been used with first grade through high school age subjects. A modification of the standard cloze procedure appears mandatory at the first grade level due to the subjects' limited writing ability. However, from second or third grade level up the cloze test appears to be an effective measure. The length of the passages used was usually 250 words, although a few investigations involved passages of more than 400 words. An optimum sample size study (Bormuth, 1965a) seems to establish an empirical basis for the selection of passage length.

Several deletion methods have been used. The most often used is a mechanical every-nth procedure in which one word in every five is replaced by a blank of 15 spaces. While this procedure appears to be entirely adequate for the purposes of measuring passage difficulty or "readability", the best method for determining the reading comprehension level of individual subjects has yet to be established. Other deletion methods have included 10 percent and 20 percent random deletions, deletions based on form class, and deletions based on a rational selection of those words which seem, in the opinion of the investigator, to be critical to the meaning transmitted.

In the vast majority of studies, the equating of passages was based on ratings by readability formula. The Dale-Chall formula was used most frequently. In several studies a publisher's estimation of passage difficulty was reported. Presentation methods seemed to be of two distinct types. The pre-cloze written method was most frequently used with large groups. However, in a few investigations an oral presentation was used in which the subject listened to the passage in its entirety prior to the cloze exercise.

Other instruments administered to validate cloze results included standardized achievement or intelligence tests and, in several instances, specially constructed multiple-choice items. These questions, based on information contained in the passages over which the cloze test was made, were carefully validated in only a few instances.

The subject population "represented" a variety of racial and socio-economic backgrounds. However, many investigators made no attempt to describe the socio-economic, racial or ethnic characteristics of the population. Frequently the bases for descriptive comments such as "representative" were not discussed. While one study involved in excess of 900 children, the great majority considered the responses of one or two hundred subjects. These were, in many instances, drawn from a single school or a single school district. The selection of the

school or district may have been predicated on administrative considerations.

The exact word-response criterion was most often used. Several investigations used alternative scoring criteria such as synonyms, form class similarity, or "grammaticality". The comparison of cloze scores with standard scores seemed to hold up well and was reported significant in virtually all studies. Most investigations with children involve methodological types of independent variables; only a few studies have used the cloze procedure as a teaching technique. The value of cloze under these conditions has yet to be demonstrated.

The summary of studies of adults. Of the 23 adult studies reported, the substantial majority are methodological investigations. Cloze scores, it has been suggested, are reliable and valid measures for the assessment of adult reading capabilities. There seems to be a substantial positive relationship between a subject's ability to use context clues, his ability to comprehend rapidly, and his general intelligence--particularly verbal intelligence. Supporting the methodological investigations involving children, the adult studies have shown that deleted words may be equally restorable at ratios greater than 1:12. However, deletions more closely spaced than five words, tend to greatly increase the difficulty of restoring the exact word deleted. Cloze scores seem to increase moderately with the decreasing density of deletions. While form class similarities seem to depend on the immediate verbal environment of a deletion, it is suggested that verbatim word predictability depends also upon the remote topical content or semantic features of the discourse in question. Cloze scores seem to correlate positively and significantly with the standardized measures used when exact-word scoring techniques are employed.

A mechanical deletion method seems to be superior to any other technique. It has been pointed out that rational deletion ceases to measure passage difficulty, since the deletions are no longer representative of possible deletions within the sample.

The relation of "sensitivity to literary style" and cloze scores has been investigated. It was indicated that word patterns of the material written in active verb form were more predictable, and presumably easier to comprehend, than those of nominalized sentences. However, the practice of using artificially-shortened sentences has been questioned, since comprehension as measured by cloze scores seems to be little effected by changes in sentence length. It is suggested that these results have implications for preparing instructional material.

For the most part, investigations on information or knowledge gain by use of cloze tests have been unsuccessful. While the early work of Taylor (1954) suggests that knowledge gain can be measured validly and

reliably with the cloze technique, later work has questioned this conclusion (Coleman, 1966). Recent studies have argued that cloze procedure is a useful technique for measuring the comprehension of oral messages under certain conditions. The need for additional investigation in this area is indicated. It has been suggested that cloze procedures may not be a sensitive device when the vocabulary and concepts contained in a passage are more advanced than the knowledge or information of the respondent would warrant.

One feature of the cloze technique which contributes to its general utility is that it can be used to construct equivalent test forms drawn from the same or similar materials. It is possible to construct multiple-test forms with similar means and variances and high intercorrelations. For the classroom teacher the cloze procedure, then, has many potential uses. The teacher can readily determine the readability of passages to be used by children with particular abilities.

It may be possible to obtain two cloze scores on the subject: (a) general comprehension skill and (b) specific comprehension relative to a particular subject. It might then be possible to examine the discrepancies between the two scores in order to provide suggestions for individualizing teaching techniques and optimizing the use of time and materials.

For the researcher, the cloze technique permits the study of some of the underlying processes involved in reading. The use of information gain or knowledge gain scores may permit the study of reading and listening in the instructional process. The study of the interrelationships between lexical and semantic comprehension might prove to be of practical value as well as theoretical value in the construction of instructional material.

CONCLUSIONS

Research to date suggests that the most valid and reliable cloze test for measuring passage difficulty is one in which:

(1) An every nth mechanical mutilation system is used.
(2) Not more than 20 words out of every 100 are deleted.
(3) Passage length is at least 250 words.
(4) Deletion ratios of 1:10 and 1:12 in longer passages may be valid for certain purposes.
(5) At least 50 words are deleted in order to insure adequate sampling of passages.
(6) The exact word deleted is indicated as the most useful and efficient scoring criteria.
(7) Other scoring systems (synonym, form class) provide less inter-scorer reliability and require substantially more time.
(8) The separate scoring of form classes or content and

function words may provide specific information for specialized purposes.

Much cloze research leaves serious questions about the generalizability of findings. Subject populations are not large and, seemingly, are selected for administrative reasons rather than by sound randomizing procedures. Information on subjects is frequently lacking in detail. This is particularly important in studies of children. In some studies data on passage difficulty and cloze test instructions are not reported. In other studies, test scores are reported on the subject's reading achievement or I.Q.

Implications of adult studies for children

In studies of adult reading and cloze research, two variables are infrequently mentioned: (a) the reading ability of the subject and (b) the difficulty of the passage (the original work of Taylor [1953, 1954, 1957] and a recent study by Miller and Coleman [1966] are noteworthy exceptions to b). It is obvious that these two variables are of importance in studies of reading and instruction of children. Nevertheless, questions on cloze methodology investigated with adult subjects are relevant to investigations with children.

For example, both Aborn, et al. (1959) and MacGinitie (1961), working with adults, reported that context of less than four words between deletions substantially reduces contextual constraint and consequently excessively reduces the discrimination power of the cloze test. These findings were used by Bormuth (1962) as a basis for the construction of his cloze materials for use with children. However, Bormuth's (1965a) data on optimum sample size in studies of children may be of questionable value when applied to adults.

Investigations using adult subjects have supported the findings of studies with children in concluding that:

(1) Consideration of synonyms and similarity of form class in scoring cloze responses or constructing cloze tests adds little if any discrimination power to the test but does raise test score means.
(2) The indiscriminate shortening of sentences has little effect on the comprehensibility of passages.

Implications for additional research

Little investigation has been reported in several fundamental areas. Much work must be done considering the optimum methodology for assessing individual reading comprehension for children. Exploration of differing deletion rates, passage lengths, and difficulty levels, in addition to separate form class scoring, should be considered. The measurement of information gain may be possible under certain

circumstances through use of the cloze test.

New areas in which the cloze procedure may be used include the evaluation of the comprehension of instructional language. This research may include both oral and written presentation. The relationships between presentation length, frequency, difficulty, and student ability may be explored with cloze techniques.

At least two components in reading comprehension may be considered; syntactic complexity and information load. In cloze research, these components have not been studied systematically. Assuming that they could be operationally defined, it may be possible to hold either or both of these variables constant in order to determine the relationships between cloze scores and these components and between the components themselves. While cloze scores may be reliable and valid measures of gross comprehension, it has yet to be shown that cloze scores are sensitive to either syntactic complexity or information load per se.

One might hypothesize, for example, that cloze scores over function word deletions are more related to syntactic complexity than to information load. On the other hand, deleted content words may have more to do with information load than syntactic complexity.

While certain significant investigations of cloze methods have contributed to their potential usefulness, much remains to be learned about the value of this relatively new technique in the study and use of language in the instructional setting.

AUTHOR INDEX

Aborn, M.....3, 4, 16, 20, 40, 41

Betts, E.A.....7

Bloomer, R.H.....16, 18, 32, 34

Blumenfeld, J.P...29, 30

Bormuth, J.R.....2, 4, 5, 7, 11, 12, 13, 14, 15, 17, 29, 31, 37, 40

Chall, J.....5, 37

Coleman, E.B.....7, 19, 20, 26, 29, 30, 32, 34 39, 40

Dale, E.....5, 37

Darnell, D.K.....29, 31, 32, 35

Deutsch, M.....8, 10

DeVito, J.A.....35, 36

Dickens, M.....32, 36

Feldman, S.....29, 30

Fillenbaum, S.....23, 25

Fletcher, J.E.....6, 7, 19, 22

Flesch, R.....5

Gallant, R.....2, 8, 9

Greene, F.P.....6, 26, 28

Hafner, L.E.....6, 23, 25

Harris, A.J.....7

Jenkinson, M.E.....5, 6

Kingston, A.J.....6, 23, 24, 29

Luke, M.....26, 27

Louthan, V.....16, 18, 26

MacGinitie, W.H.....4, 19, 22, 40

McLeod, J.....11, 12, 26

Miller, G.A.....26

Miller, G.R.....19, 20, 40

Musgrave, B.....26, 27

Osgood, C.E.....2

Portnoy, S.....29, 30

Rankin, E.F.....4, 6, 29

Rubenstein, H.....3, 16

Ruddell, P.B.....2, 4, 6, 11, 12

Salzinger, K.....29, 30, 35

Schneyer, J.W.....14, 17

Selfridge, J.....26

Spache, G.....8

Sterling, T.D.....3, 16

Strickland, R.G.....12

Taylor, W.L.....1, 2, 3, 4, 5,

Taylor, W.L.(Cont.)..7, 19, 21, 32,
33, 35, 38, 40

Thorndike, E.L.....7

Weaver, W.W.....6, 23, 24, 29

Williams, F.....32, 36

REFERENCES

Aborn, M., Rubenstein, H., & Sterling, T.D. Sources of contextual constraint upon words in sentences. *Journal of Experimental Psychology*, 1959, 57 (3), 171.

Bennett, S., Semmel, M.I., & Barrett, L.S. In Lane, H.L. & Zale, E.M. (Eds.) Studies in language and language behavior. Center for Research on Language and Language Behavior, University of Michigan, 1965, 414-428.

Betts, E.A. *Foundations of reading instruction.* New York: American Book Company, 1946.

Bloomer, R.H. The cloze procedure as a remedial reading exercise. *Journal of Developmental Reading*, 1962, 5, 173-181.

Bloomer, R.H. & Heitzman, A.J. Pretesting and the efficiency of paragraph reading. *Journal of Reading*, 1965, 8 (4), 219.

Bloomer, R.H., Louthan, V., & Heitzman, A.J. Non-overt reinforced cloze procedure. U.S. Office of Education Project Report #2245, University of Connecticut, 1966.

Blumenfeld, J.P. & Miller, G.R. Improving reading through teaching grammatical constraints. *Elementary English*, 1966, 43, 752.

Bormuth, J.R. Cloze tests as measures of readability and comprehension ability. Unpublished doctoral dissertation, University of Indiana, 1962.

Bormuth, J.R. Experimental applications of cloze tests. *International Reading Association Conference Proceedings*, 1964, 9, 303. (a)

Bormuth, J.R. Mean word depth as a predictor of comprehension difficulty. *California Journal of Educational Research*, 1964, 15 (5), 226. (b)

Bormuth, J.R. Relationships between selected language variables and comprehension ability and difficulty. Cooperative Research Project #2082, U.S. Office of Education, 1964. (c)

Bormuth, J.R. Optimum sample size and cloze test length in readability measurement. *Journal of Educational Measurement*, 1965, 2 (1), 111. (a)

Bormuth, J.R. Validities of grammatical and semantic classifications of cloze scores. *International Reading Association Conference Proceedings*, 1965, 10, 283. (b)

Bormuth, J.R. & MacDonald, O.L. Cloze tests as a measure of ability to detect literary style. In J. Allen Figurel (Ed.), *Reading and Inquiry, International Reading Association Conference Proceedings*, 1965, 10, 287-290. (c)

Bormuth, J.R. Readability: A new approach. *Reading Research Quarterly*, 1966, 1 (3), 79. (a)

Bormuth, J.R. Design of readability research. *International Reading Association Conference Proceedings*, 1966, 11, 485-489. (b)

Bormuth, J.R. Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, 1967, 10, 291. (a)

Bormuth, J.R. The implications and use of cloze procedure in the evaluation of instructional programs. Center for the Evaluation of Instructional Programs, University of California, Los Angeles, 1967. (b)

Bormuth, J.R. Cloze readability: Criterion reference scores. *Journal of Educational Measurement*, 1968 (in press).

Carterette, E.C. & Jones, M.H. Redundancy in children's tests. *Science*, 1963, 140, 1309-1311.

Coleman, E.B. Improving comprehension by shortening sentences. *Journal of Applied Psychology*, 1962, 46, 131-134.

Coleman, E.B. & Blumenfeld, J.P. Cloze scores of nominalizations and their grammatical transformations using active verbs. *Psychological Reports*, 1963, 13, 651-654.

Coleman, E.B. Developing a technology of written instruction: Some determiners of the complexity of prose. Paper read at symposium on Verbal Learning and Written Instruction, New York, March, 1966.

Coleman, E.B. & Miller, G.R. A measure of information gained during prose learning. *Quarterly Journal of Reading*, in press.

Dale, E. & Chall, J. A formula for predicting readability. *Educational Research Bulletin*, January, 1948, 11-28.

Dale, E. & Seels, B. Readability and reading. *International Reading Association Conference Proceedings*, 1966.

Darnell, D.K. The relation between sentence order and comprehension. *Speech Monographs*, 1963, 30, 97.

Deutsch, M., Cherry, E., Maliver, H., & Brown, R. Communication for information in the elementary classroom. Institute for Developmental

Studies, New York University, 1964.

DeVito, J.A. Comprehension factors in oral and written discourse of skilled communicators. *Speech Monographs*, 1965, 32, 124.

Dickens, M. & Williams, F. An experimental application of cloze procedure and attitude measures to listening comprehension. *Speech Monographs*, 1964, 31, 103-108.

Ervin, S.M. Changes with age in the verbal determinants of word association. *American Journal of Psychology*, 1961, 74, 361-372.

Epstein, W. The influence of syntactical structure on learning. *American Journal of Psychology*, 1961, 74, 80-85.

Fillenbaum, S., Jones, L.V., & Rapoport, A. The predictability of words and their grammatical classes as a function of rate of deletion from a speech transcript. *Journal of Verbal Learning and Verbal Behavior*, 1963, 2, 186-194.

Fletcher, J.E. A study of the relationships between ability to use context as an aid in reading and other verbal abilities. Unpublished doctoral dissertation, University of Washington, 1959.

Flesch, R. *The art of readable writing*. Harper & Row, 1949.

Gallant, R. Use of cloze tests as a measure of readability in the primary grades. Proceedings of the International Reading Association Convention, 1965, 10, 286-287.

Greene, F.P. A modified cloze procedure for assessing adult reading comprehension. Unpublished doctoral dissertation, University of Michigan, 1964.

Greene, F.P. Modification of the cloze procedure and changes in reading test performances. *Journal of Educational Measurement*, 1965, 2 (2), 213-217.

Hafner, L.E. Relationships of various measures of the cloze. In E.L. Thurston & L.E. Hafner (Eds.) *Thirteenth Yearbook of the National Reading Conference*. Milwaukee, Wisconsin, National Reading Conference, Inc., 1963, 135-145.

Hafner, L.E. Implications of cloze. In E.L. Thurston & L.E. Hafner (Eds.) *The philosophical and sociological bases of reading*. Fourteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin, The National Reading Conference, Inc., 1965, 151-158.

Hafner, L.E. Cloze procedure. *Journal of Reading*, 1966, 9, 415-421.

Harris, A.J. *Effective teaching of reading.* New York: David McKay, 1962.

Jenkinson, M.E. Selected processes and difficulties in reading comprehension. Unpublished doctoral dissertation, University of Chicago, 1957.

Kerfoot, J.F. Reading in the elementary school. *Review of Educational Research,* 1967, 37, 120.

Kingston, A.J. & Weaver, W.W. Recent development in readability appraisal. *Journal of Reading,* 1967, 11, 44.

Klare, G.R. *The measurement of readability.* Ames, Iowa: Iowa State University Press, 1963.

Lee, W.D. What does research in readability tell the classroom teacher? *Journal of Reading,* 1964, 8, 141-144.

Louthan, V. Some systematic grammatical deletions and their effects on reading comprehension. *English Journal,* 1965, 54, 295.

Luke, M. Form class and cloze procedure. Unpublished manuscript, Language Development Program, Center for Human Growth and Development, University of Michigan, 1964.

MacGinitie, W.H. Contextual constraint in English prose paragraphs. *Journal of Psychology,* 1961, 51, 121-130.

Marks, M.R. & Taylor, W.L. The influence of contextual and goal constraints on the meaningfulness of "automatic sentences". *Journal of Social Psychology,* 1954, 40, 43-51.

McLeod, J. *GAP reading comprehension test.* Melbourne, Australia: Heinemann, 1965.

McLeod, J. & Anderson, J. Readability assessment and word redundancy of printed English. *Psychological Reports,* 1966, 18, 35-38.

Miller, G.A. & Friedman, E.A. The reconstruction of mutilated English texts. *Information and Control,* 1957, 1, 38-55.

Miller, G.A. & Selfridge, J. Verbal context and the recall of meaningful material. *American Journal of Psychology,* 1950, 63, 176-185.

Miller, G.R. & Coleman, E.B. A set of 36 passages calibrated for comprehensibility. Inglewood, California: Southwest Regional Laboratory for Educational Research and Development, 1966.

Musgrave, B. The effect of who and what context on cloze and

commonality scores. *Journal of Social Psychology*, 1963, 59, 185-192.

Osgood, C.E. The nature and measurement of meaning. *Psychological Bulletin*, 1962, 49 (3), 197-237.

Rankin, E.F. An evaluation of the cloze procedure as a technique for measuring reading comprehension. Unpublished doctoral dissertation, University of Michigan, 1957.

Rankin, E.F. Uses of the cloze procedure in the reading clinic. In J.A. Figurel (Ed.), IRA Conference Proceedings, New York. *Scholastic Magazine*, 1959, 4, 228-232.

Rankin, E.F. The cloze procedure--Its validity and utility. In O.S. Causey and W. Eller (Eds.), *Eighth Yearbook of the National Reading Conference*, National Reading Conference, Inc., 1959, 8, 131-144.

Rankin, E.F. The cloze procedure--A survey of research. In E.L. Thurston and L.E. Hafner (Eds.) *The philosophical and sociological bases of reading*. Fourteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin, The National Reading Conference, Inc., 1965, 133-150.

Rankin, E.F. Residual gain as a measure of individual differences in reading improvement. *Journal of Reading*, 1965, 10, 224-233.

Rankin, E.F. Research design and the cloze procedure. Proceedings of the International Reading Association Convention, 1966, 2 (1), 489-491.

Ruddell, R.B. The effect of oral and written patterns of language structure on reading comprehension. Unpublished doctoral dissertation, University of Indiana, 1963.

Ruddell, R.B. A study of the cloze comprehension technique in relation to structurally controlled reading material. In J.A. Figurel (Ed.) *Improvement of reading through classroom practice*. International Reading Association Conference Proceedings, 1964, 9, 298-306.

Ruddell, R.B. The effect of oral and written patterns of language structure on reading comprehension. *Reading Teacher*, 1965, 18, 270.

Salzinger, K., Portnoy, S., & Feldman, R.S. The effect of order of approximation to the statistical structure of English on the emission of verbal responses. *Journal of Experimental Psychology*, 1962, 64, 52-57.

Schneyer, J.W. Use of the cloze procedure for improving reading comprehension. *Reading Teacher*, 1965, 19, 174-179.

Spache, G. A new readability formula for primary grade materials. *Elementary School Journal*, 1952, 53 (7), 410-413.

Strickland, R.G. The language of elementary school children: Its relationship to the language of reading textbooks and the quality of reading of selected children. *Bulletin of the School of Education*, University of Indiana, 1962, 38 (4).

Taylor, W.L. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 1953, 30, 414-438.

Taylor, W.L. Application of "cloze" and entropy measures to the study of contextual constraint in samples of continuous prose. Unpublished doctoral dissertation, University of Illinois, 1954.

Taylor, W.L. Recent developments in the use of the cloze procedure. *Journalism Quarterly*, 1956, 33, 42-48.

Taylor, W.L. Cloze readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology*, 1957, 41, 19-26.

Thompson, E. Some effects of message structure on listeners comprehension. *Speech Monographs*, 1957, 24, 51.

Thorndike, E.L. Reading and reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology*, 1917, 8, 323-332.

Tomlinson, I.N. Sources of contextual constraint upon words in sentences. Unpublished masters thesis, University of California, Los Angeles, 1967.

Weaver, W.W. The predictibility of omissions in reading and listening. *Eleventh Yearbook of the National Reading Conference*, Milwaukee, Wisconsin, 1962, 148-153.

Weaver, W.W. Theoretical aspects of the cloze procedure. In E.L. Thurston and L.E. Hafner (Eds.) *The philosophical and sociological bases of reading*. Fourteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin, The National Reading Conference, 1965, 115-132.

Weaver, W.W. & Kingston, A.J. A factor analysis of the cloze procedure and other measures of reading and language ability. *Journal of Communication*, 1963, 13, 252-261.